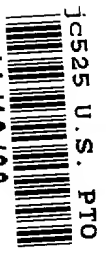


A

11/12/98



IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

Inventor.....Sarah E. Zuberec et al.
Applicant.....Microsoft Corporation
Attorney's Docket No.MS1-286US
Title: Speech Recognition User Interface



TRANSMITTAL LETTER AND CERTIFICATE OF MAILING

To: Commissioner of Patents and Trademarks
Washington, D.C. 20231
From: Daniel L. Hayes (509) 324-9256
Lee & Hayes, PLLC
W. 201 North River Drive, Suite 430
Spokane, WA 99201

The following enumerated items accompany this transmittal letter and are being submitted for the matter identified in the above caption.

- 1. Transmittal Letter with Certificate of Mailing included.
- 2. PTO Return Postcard Receipt
- 3. New patent application (title page plus 30 pages, including claims 1-39 & Abstract)
- 4. Executed Declaration
- 5. 7 sheets of formal drawings (Figs. 1-7c)
- 6. Assignment w/Recordation Cover Sheet

Large Entity Status ☒ Small Entity Status ☐

The Commissioner is hereby authorized to charge payment of fees or credit overpayments to Deposit Account No. 50-0463 in connection with any patent application filing fees under 37 CFR 1.16, and any processing fees under 37 CFR 1.17.

Date: 11/12/98
By: Daniel L. Hayes
Reg. No. 34,618

CERTIFICATE OF MAILING

I hereby certify that the items listed above as enclosed are being deposited with the U.S. Postal Service as either first class mail, or Express Mail if the blank for Express Mail No. is completed below, in an envelope addressed to The Commissioner of Patents and Trademarks, Washington, D.C. 20231, on the below-indicated date. Any Express Mail No. has also been marked on the listed items.

Express Mail No. (if applicable) EL206425397

Date: 11/12/98
By: Dana L. Calhoun

E1206425397

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

APPLICATION FOR LETTERS PATENT

Speech Recognition User Interface

Inventor(s):

Sarah E. Zuberec

Cynthia DuVal

Benjamin N. Rabelos

ATTORNEY'S DOCKET NO. MS1-286US

1 **TECHNICAL FIELD**

2 This invention relates to speech recognition systems. More particularly,
3 this invention relates to user interfaces for speech recognition systems. This
4 invention further relates to vehicle computer systems and other electronic devices
5 that implement such speech recognition user interfaces. This invention further
6 relates to methods and systems for facilitating communication between people
7 through a device.

8
9 **BACKGROUND**

10 Two common types of speech recognition systems are continuous and
11 discrete. Continuous speech recognition systems detect and discern useful
12 information from continuous speech patterns. In use, an operator may speak
13 phrases and sentences without pausing and the continuous speech recognition
14 system will determine the words being spoken. Continuous speech recognition
15 systems are used, for example, in voice-input word processors that enable
16 operators to dictate letters directly to the computer.

17 Discrete speech recognition systems are designed to detect individual
18 words and phrases that are interrupted by intentional pauses, resulting in an
19 absence of speech between the words and phrases. Discrete speech recognition
20 systems are often used in "command and control" applications in which an
21 operator speaks individual commands to initiate corresponding predefined control
22 functions. In a typical use, the operator speaks a command, pauses while the
23 system processes and responds to the command, and then speaks another
24 command. The system detects each command and performs the associated
25 function.

1 In all speech recognition systems, various forms of feedback are used to
2 indicate to the user when the system is active and ready for speech input. In many
3 PC based systems, feedback is provided by means of onscreen visual elements.
4 As an example, in some commercially available dictation systems, an icon is
5 present and flashing indicating to the user that he/she can begin dictation. Text
6 appears on screen as spoken words begin to be recognized. In this case, users are
7 trained that they can speak at any time until they actively shut the recognition
8 system off.

9 In data access systems, feedback is provided by spoken or audio prompts.
10 As an example, feedback can be modeled after a conversation. The system speaks
11 a key word or phrase, followed by a pause. It is after this pause that the user must
12 respond with their chosen command. In this example, users are trained that they
13 must speak after the pause and before the system times out.

14 Not all environments that employ a speech recognition system have the
15 luxury of providing such clean exchanges between the system and user (i.e.,
16 knowing when the system speaks and when the user speaks). In some
17 environments, users are concentrating on a primary task and using speech as a
18 method of input because their hands and eyes are otherwise occupied. In this
19 situation, feedback needs to be quick and succinct, requiring little attention from
20 the user.

21 Speech interface systems can be designed to be always awake and available
22 to accept speech commands from the user. This is very much like how two people
23 hold a conversation. Even if one person is talking, they can still hear responses
24 from the other person. Both talking and listening can be done at the same time.
25 While this is a natural style of interaction, technical limitations of certain speech

1 systems do not allow it. In many cases, if the system is always awake, it may
 2 recognize any extraneous sound it hears. For instance, if a speech system in a car
 3 is always listening for all speech commands while the radio is playing, the system
 4 may pick up words from the radio and carry out actions not intended by the
 5 vehicle operator. This is confusing and frustrating for the operator.

6 To avoid this potentially confusing situation, speech systems can be
 7 designed to be awake for limited periods of time and when awake, to utilize
 8 limited sets of recognizable words. A complete list of recognized words or
 9 phrases is referred to as the "vocabulary", and a subset of the vocabulary that the
 10 recognition system is attempting to detect at any one time is known as the
 11 "grammar." In general, the smaller the active grammar, the more reliable the
 12 recognition because the system is only focusing on a few words or phrases.
 13 Conversely, the larger the active grammar, the less reliable the recognition because
 14 the system is attempting to discern a word or phrase from many words or phrases.

15 Once a command is given and accepted by the system, the user is given a
 16 predefined time limit to speak other commands in the grammar before the system
 17 goes back to sleep and stops accepting commands. Since the system is initially
 18 listening for only one or two commands, random and unwanted recognition of
 19 extraneous words is greatly reduced. However, operating a speech system that has
 20 sleep and active modes, as well as changing grammars, can be difficult and/or
 21 confusing to the operator in the absence of some form of feedback.

22 Accordingly, there is a need for speech recognition user interfaces to guide
 23 operators through the various states and options of the speech system by using
 24 cues that can be readily ascertained by the user in a casual hands-free, at a glance
 25 environment.

Another problem contemplated by the inventors concerns other types of communications that rely on asynchronous messages. For example, video conferencing, teleconferencing, and certain network-based software provide a distributed collaboration environment in which two or more people collaborate. In such situations, it is difficult to tell in some remote collaboration software that a person at one of site has tried to break into the conversation.

In this distributed collaboration environment, common face-to-face cues that people intuitively rely on to know when to enter into the conversation may not be available. The video may be blurry, or not all participants are visible, or other problems may prevent traditional conversational cues.

Accordingly, there is a need for a system that provides visual and/or auditory cues to facilitate distributed communications where traditional visual feedback is unattainable for technical and other reasons.

SUMMARY

This invention concerns a speech recognition system having a user interface that provides both visual and auditory feedback to guide a user in a casual, eyes-off environment. The speech recognition (SR) system includes an application, a vocabulary accessible by the application that holds a set of utterances applicable to the application, and a speech recognition engine to recognize the utterances in the vocabulary. The user interface provides feedback to inform a user when the speech recognition engine is awaiting vocal input and to confirm when an utterance is recognized.

The user interface includes an audio sound or speech generator that produces three distinct sounds: a SR "on" sound signifying that the system is on

1 and actively awaiting vocal input; an "off" sound indicating that the SR system is
2 off and in a sleep mode; and a "confirm" sound noting that an utterance has been
3 recognized. The "on" sound is triggered by a key "wake up" command or by
4 depression of button. Once awake, the speech recognition engine expects to
5 receive an utterance within a predetermined response time. The "confirm" sound
6 signals the start of the response time. If the response time lapses before a
7 recognizable utterance is entered, the "off" sound is played.

8 The user interface further includes a visual component in the form of a
9 graphic that changes with the tolling of the response period. In one
10 implementation, the count graphic is a progress bar that counts down or shortens
11 in proportion to the diminishment of the response period. When the response time
12 runs out, the progress bar disappears entirely. On the other hand, if the speech
13 engine recognizes an utterance within the response period, the user interface plays
14 the "confirm" sound and restarts the countdown graphic. The user interface may
15 also temporarily change the color of the graphic elements from one color to
16 another and then back to the original color to reflect a correct voice entry.

17 The user interface may also be used in distributed collaboration efforts to
18 facilitate conversations between remotely located people. The visual display tells
19 a user when they can speak and how long they can speak before their turn "times
20 out."

21 22 **BRIEF DESCRIPTION OF THE DRAWINGS**

23 Fig. 1 is a functional block diagram of a discrete speech recognition system.

24 Fig. 2 is a diagrammatic illustration of a vehicle computer system that
25 implements the discrete speech recognition system.

Fig. 3 is a block diagram of the vehicle computer system of Fig. 2.

Fig. 4 is a block diagram of a more general computer device used to implement the speech recognition system.

Fig. 5 is a flow diagram showing steps in a method for operating a speech recognition system.

Fig. 6 illustrates how a default grammar is expanded multiple times to create multiple ballooned grammars during operation.

Figs. 7a-7c illustrates a graphical interface having a countdown graphic to provide visual feedback pertaining to response duration.

DETAILED DESCRIPTION

This invention concerns speech recognition systems and user interfaces for such systems. The invention is described in the context of a discrete speech recognition system. However, aspects of the invention can be applied to continuous speech recognition systems as well. Moreover, this invention may be utilized in a distributed collaboration system (e.g., video conferencing, teleconferencing, collaboration software, etc.).

Fig. 1 shows a discrete speech recognition system 20 that recognizes individual words or phrases. As used herein, the term “utterances” means a word, phrase, or other spoken sound that is detectable as a voice input to the speech recognition system. The discrete speech recognition system 20 includes an application 22, a vocabulary 24, an active grammar 26, a speech recognition engine 28, and a user interface 30. The application 22 may be essentially any kind of application that supports voice-input commands for operation. One suitable type of application is the command and control application where discrete

commands are issued to cause a particular function to occur. As possible examples, the application 22 might be a program to operate a vehicle radio, or a program to locate an address or destination, or a program to operate a telephone, or a program to open files on a computer, and so forth.

The vocabulary 24 is the complete list of utterances that are recognized by the application 22. The vocabulary 24 is stored in memory that can be accessed by the application. The active grammar 26 is a subset of the vocabulary that the recognition system is attempting to recognize at any one time. During operation, the application 22 extracts various sets of utterances from vocabulary 24 and registers them as the active grammar. The active grammar 26 is stored in short-term memory or cache, which is accessible by the application 22 and the speech recognition engine 26.

The speech recognition engine 28 performs the actual task of recognizing utterances in the active grammar. It compares the voice input from the operator with utterances in the active grammar 26. Upon detection of an utterance listed in the active grammar 26, the speech recognition engine 28 informs the application 22 of the particular utterance that has been recognized. The application 22 then performs the function associated with the utterance.

The user interface 30 combines audible and visual feedback to inform a user when the system 20 is expecting a voice command. The audible feedback is in the form of different tones that indicate different response actions on the part of the system 20. The visual feedback provides at-a-glance interpretation of how long the system will or has been listening for a voice command. The user interface 30 improves user interaction with the speech recognition system 20 by

1 conveying to the user that the system is actively listening for a recognizable
2 utterance or has heard and understood a recognizable utterance.

3 The speech recognition system 20 begins with a starter word to initially
4 activate the system for further voice inputs. Using a starter word helps narrow
5 initiation of the system to only those instances that the user actually wants to open
6 a dialog with the system. Other words spoken by the user or by others in
7 background (e.g., words from a radio or TV broadcast, or from a passerby, etc.)
8 will rarely trigger initiation of the system. For instance, suppose the starter word
9 is "WAKE". In the absence of the "WAKE" command, the system 20 remains in
10 its sleep or dormant state.

11 Alternatively, the system may be activated through a hardware control. The
12 device may include a "Push to Talk" or "PPT" button that, upon depression, wakes
13 up the system.

14 When it recognizes the command "WAKE" or detects actuation of the Push
15 to Talk button, the speech recognition system 20 awakes to its active state and
16 becomes ready to receive further audible input commands. The user interface 30
17 emits a first tone or other sound signifying to the user that the system recognizes
18 the wake up command. Concurrently, the user interface 30 displays an onscreen
19 countdown graphic to provide additional visual feedback that the system is awake.

20 Once active, the speech recognition system 20 waits for a predetermined
21 duration for another recognizable command. The countdown graphic visually
22 decrements with passage of time. If the user chooses to glance at the display, the
23 countdown graphic informs the user quickly as to approximately how much time is
24 left to say a command.
25

Each time the user speaks a word that is recognized, the user interface 30 plays a second tone or sound to confirm recognition of the word and restarts the countdown graphic. If the user stops providing commands and the predetermined duration lapses, the countdown graphic visually decrements until it disappears entirely and the user interface sounds a third tone indicating that the response time has elapsed. The speech recognition system returns to its dormant state and begins listening for the starter word.

The discrete speech recognition system 20 can be implemented in many different contexts. Three exemplary contexts are provided in this disclosure: an vehicle computer/entertainment accessory, a standalone computing device, and networked computing devices. These implementations are discussed separately below.

Exemplary Context 1: Vehicle Accessory

Fig. 2 shows an in-dash vehicle accessory 50 that implements the discrete speech recognition system 20. In the illustrated implementation, the vehicle accessory 50 is a vehicle computer system that includes a variety of different elements such as a security subsystem, a navigation subsystem, and an entertainment subsystem. In other embodiments, the accessory might be a car radio, cassette player, CD player, or similar device. The speech recognition system facilitates voice activation of the accessory 50, or features thereof, to permit the driver to verbally enter commands in a hands-free, eyes-free environment.

Vehicle accessory 50 has a base module 52 and a detachable faceplate 54. Base module 52 includes a housing 56, which is sized to be mounted in a vehicle

1 dashboard similar to a conventional car stereo. Preferably, housing 56 has a form
2 factor of a single DIN (Deutsche Industry Normen). The base module could
3 alternatively be housed in a 2 DIN unit or other special form factor for an OEM.

4 In the described implementation, base module 52 is a computer that runs a
5 multi-tasking operating system to support multiple applications. One preferred
6 operating system is a Windows brand operating system sold by Microsoft
7 Corporation, such as Windows 95, Windows CE, Windows NT or other derivative
8 versions of the Windows family of operating systems. The computer and
9 operating system are configured to support after-market peripherals including both
10 hardware and software components.

11 The computer 52 includes at least one storage drive that permits the vehicle
12 user to download programs and data from a storage medium. In the illustrated
13 implementation, the computer 52 has a CD ROM drive 58 that reads application-
14 related CDs, as well as musical, video, game, or other types of entertainment CDs.
15 In this manner, the CD ROM drive 58 performs a dual role of storage drive and
16 entertainment player. Also, a hard disk drive (not shown in Fig. 4) that can be
17 used for storing both application programs and user data is included on the
18 computer module 52. The computer base module 52 has an optional 3.5" floppy
19 diskette drive 60, a smart card reader 62, and dual PCMCIA card sockets 64 which
20 accept PC (or PCMCIA) card types II and III. The faceplate 54 can be rotated to
21 expose and permit easy access to the storage drives. In other implementations, the
22 faceplate 54 has slots to permit access to one or more of the memory drives.

23 Faceplate 54 has a keypad 66 and a graphical display 68. The display 68 is
24 preferably a backlit LCD having a rectangular array of pixels that are individually
25 selectable for illumination or display. The display element is a medium-

1 resolution, bit-mapped display system having at least 10,000 pixels. In the
2 described implementation, the array of pixels has a size of at least 64 by 256
3 pixels. The operating system of base module 52 interacts with faceplate keypad
4 66 and faceplate display 68 as peripheral devices when the faceplate 54 is attached
5 to the base module 52.

6 The faceplate 54 has a start button 70 that provides the familiar “Start”
7 functions of a Windows brand operating system. The faceplate 54 also has a
8 power button 72, a four-position actuator 74, an “enter” button 76, an “escape”
9 button 78, a “help” button 80, a “menu” button 82, and a volume control input 84.

10 The accessory 50 has a voice input port 86 that picks up voice commands
11 for entry to the speech recognition system. The accessory also supports an IrDA
12 (infrared developers association) transceiver port 88 mounted on the faceplate 54
13 to transmit and receive data and programs using infrared signals.

14 The vehicle accessory 50 can be used to integrate multiple vehicle-related
15 systems onto one open platform. For instance, the accessory can serve as a
16 multimedia entertainment system, a navigation system, a communications system,
17 a security system, and a diagnostics system. Moreover, the accessory provides
18 additional functionality traditionally associated with desktop and laptop personal
19 computers. For instance, it can support word processing applications, spreadsheet
20 applications, database applications, and appointment/schedule applications.
21 Furthermore, the vehicle accessory 50 can be configured to operate as a server to
22 other computing units in a vehicle to distribute games, video movies, and the like
23 to passengers.

24 Vehicle accessory 50 may support a number of components and features
25 that are not discussed herein. For more detail regarding such features refer to a

1 U.S. Patent 5,794,164, entitled "Vehicle Computer System", which was issued
2 August 11, 1998 and is assigned to Microsoft Corporation. This patent is hereby
3 incorporated by reference.

4 Fig. 3 is a simplified block diagram showing pertinent components of
5 vehicle accessory 50. It has a conventional computer architecture, including a
6 microprocessor 100, associated addressable memory 102, peripheral drivers 104,
7 and optional mass storage 106. In addition, accessory 50 includes keypad 66, a
8 graphics display 68, a voice input port 86, and an audio output 110 to permit
9 emission of sound. The various components communicate internally through a
10 parallel data bus 112. The audio output can be configured as a sound or tone
11 generator integrated into the stereo functionality, or be.

12 The addressable memory 102 is shown to include high-speed volatile
13 memory 114 and non-volatile memory 116. An operating system 118 and one or
14 more programs 120 are stored in memory 116 and execute on the microprocessor
15 100. Alternatively, the operating system and programs can be stored in mass
16 storage 106 (as represented by the dashed boxes). Other components, not shown,
17 provide different types of entertainment functionality as described in the patent
18 cited above.

19 In the illustrated implementation, the speech recognition system 20 is
20 embodied as a computer program or callable modules that are stored in memory
21 116 and executed on microprocessor 100. Voice commands are entered into the
22 vehicle computer via the voice input 86. The speech recognition system 20
23 receives voice commands and performs functions based upon recognized
24 commands. Visual and audible feedback forming the user interface 30 are
25 implemented by the display 68 and audio output 110.

1 The data processor 100 is programmed by means of instructions stored at
 2 different times in the various computer-readable storage media of the device.
 3 Programs are typically distributed, for example, on floppy disks or CD-ROMs.
 4 From there, they are installed or loaded into the secondary memory of a computer.
 5 At execution, they are loaded at least partially into the computer's primary
 6 electronic memory. The invention described herein includes these and other
 7 various types of computer-readable storage media when such media contain
 8 instructions or programs for implementing the steps and features described below
 9 in conjunction with a microprocessor or other data processor. The invention also
 10 includes the computer and other devices themselves when programmed according
 11 to the methods and techniques described below.

12 Vehicle accessory 50 can be used to integrate multiple vehicle-related
 13 systems onto one open platform. For instance, the system can serve as a
 14 multimedia entertainment system, a navigation system, a communications system,
 15 a security system, and a diagnostics system. Moreover, the system provides
 16 additional functionality traditionally associated with desktop and laptop personal
 17 computers. For instance, with the addition of a peripheral display device it can
 18 support word processing applications, spreadsheet applications, database
 19 applications, and appointment/schedule applications. Furthermore, the vehicle
 20 computer system 50 can be configured to operate as a server to other computing
 21 units in a vehicle to distribute games, video movies, and the like to passengers.

22 **Exemplary Context 2: Computing Device**

23 The speech recognition system 20 may be implemented in devices other
 24 than a vehicle accessory. To provide another example for discussion purposes,
 25

Fig. 4 is a simplified block diagram of a computing device 150 that implements the discrete speech recognition system (DSRS) 20. The computing device 150 has a processor 152, associated addressable memory 154 (including non-volatile memory and high-speed volatile memory), a voice input device 156, a sound generator 158, and a display 160. The speech recognition system 20 is embodied as a software or firmware program stored or embedded in memory 154 and executes on the processor 152. The speech recognition program 20 has an application 22, a predefined vocabulary 24, an active grammar 26, a recognition engine module 28, and a user interface 30.

The computer device 150 is representative of many diverse products. Examples of representative products include pagers, cellular phones, handheld personal information devices, stereos, VCRs, set-top boxes, calculators, appliances, and so forth. It may also represent computers such as desktop computers, laptop computers, handheld computers, and other types of computer-based products.

Exemplary Context 3: Networked Computing Devices

The speech recognition system 20 may be implemented in multiple devices that are interconnected via a network (e.g., telephone, cable, Internet, LAN, WAN, etc.). Each device can be implemented as shown in Fig. 4, but adapted with a network port to facilitate communication with other devices over a network. This architecture forms a collaboration system in which multiple users are allowed to collaborate from remote locations. This architecture is representative of teleconferencing, video conferencing, and other collaboration-centric systems.

1 The speech recognition system is implemented at each node. The system
2 provides visual and auditory feedback to help a person know when it is his/her
3 turn to speak and when his/her turn is about to "time out".

4 5 **Operation of Speech Recognition User Interface**

6 Fig. 5 shows steps in a method for operating the speech recognition system
7 in a manner that improves human interaction with the system. The process is
8 described with additional reference to an exemplary grammar context illustrated of
9 Fig. 6 and an exemplary graphical interface shown in Figs. 7a-7d.

10 At step 200 in Fig. 4, the speech recognition system 20 continually listens
11 for a starter command, such as "WAKE". Essentially, the active grammar at this
12 point is loaded only with the starter word "WAKE" during the sleep mode, as
13 indicated by grammar 250 in Fig. 6. Until it hears the starter word "WAKE" (i.e.,
14 the "no" branch from step 200), the system 20 remains in sleep mode with
15 sufficient resources left operating to listen for the starter word. The user may
16 wake up the system at any time by speaking the starter word "WAKE" into the
17 voice input port 86 (or 156).

18 Upon speaking the starter word "WAKE", the speech recognition engine 28
19 recognizes the starter word (i.e., the "yes" branch from step 200) and the user
20 interface emits a distinctive "on" tone or sound from the audio output 110 or
21 sound generator 158 (step 202). The audio feedback immediately indicates to the
22 user that the system heard and understood the starter command and is now awake
23 and actively listening for the next command. The audible feedback may be in the
24 form of one or more tones, or a spoken reply like "Computer is Active".
25

1 The speech recognition system 20 starts a response time in which it expects
2 to hear a next command from the user (step 204 in Fig. 5). As an example, the
3 response time might be ten seconds, meaning that the user has ten seconds from
4 the time of the first tone to speak the next command. The response time is
5 configurable, allowing the user to set it to any desired duration. The speech
6 recognition user interface 30 displays an onscreen countdown graphic to provide
7 additional visual feedback that the system is awake (step 206 in Fig. 5). The
8 countdown graphic visually changes with passage of time so that the user quickly
9 ascertains how much time is left to respond.

10 Fig. 7a shows a graphical interface 300 illustrating one exemplary
11 implementation of the countdown graphic. The visual interface 300 is shown
12 displayed on the vehicle computer screen 68 (or device display 160). In this
13 example, the graphical user interface 300 exposes an address book application, as
14 indicated by the address card icon 302. The address book application organizes
15 and maintains personal information (e.g., address, telephone number, fax number,
16 email address, etc.). Here, the user interface 300 lists the name "Alvin Adams".
17 Other available applications are represented by icons 304. The graphical user
18 interface 300 further includes a navigation glyph 306 at one end of the screen to
19 assist the user with navigating the interface.

20 A countdown graphic 310 is formed along the bottom of the interface 300
21 as an elongated progress bar. As time runs out, the countdown graphic 310
22 shortens in appearance proportionately to the amount of time that has passed. Fig.
23 7b shows the countdown graph 310 after approximately three seconds has elapsed
24 from a previous tone. Fig. 7c shows the countdown graph 310 after approximately
25 eight seconds has elapsed from a previous tone.

At step 208 in Fig. 5, the application 22 in the speech recognition system 20 extracts a default grammar from the vocabulary 24 and loads the default grammar into the active grammar 26. For purpose of continuing discussion, suppose the discrete speech recognition system 20 is implemented to control an address book program implemented by the vehicle computer/entertainment accessory 50 or the computing device 150. It is noted that the speech recognition system can be used to control many other functions in addition to an address book. Other controllable functions include operating a radio or stereo parameters, adjusting temperature control, operating a cellular phone, operating a CD player, locking and unlocking doors, lowering and raising windows, and so forth.

The default active grammar, referenced by number 252 in Fig. 6, contains four utterances: "Place", "Name", "Directions", and "Address". Table 1 contains the default grammar and associated functions.

Table 1: Default Grammar 252

<u>Word/Phrase</u>	<u>Command Function</u>
Place	Keyword to place menu
Name	Sets address book to specified name.
Directions	Keyword to Directions menu
Address	Keyword to address menu

The first, third, and fourth utterances—"Place", "Directions", and "Address"—are keyword utterances and the second—"Name"—is a non-keyword utterance. Keyword utterances are used to balloon the active grammar to include utterances in addition to the default grammar.

At step 210 in Fig. 5, the speech recognition engine 28 begins listening to speech patterns in an effort to detect an utterance contained in the default grammar 252. The recognition engine 28 continues to listen for the preset response time (i.e., ten seconds), as indicated by the loop containing the "no" branch from step 210, the decision step 212 and the "no" branch from step 212. If the recognition engine 28 fails to hear an utterance from the default grammar (or any current grammar) within the prescribed response time (i.e., the "no" branch from step 210 and the "yes" branch from step 212 in Fig. 5), speech recognition system emits distinctive "off" tone or sound from the audio output 110 or sound generator 158 (step 228). This audible sound indicates that the system is returning to its sleep mode and will require restatement of the starter word to awaken the device. Concurrently, the countdown graphic disappears to provide a visual indication that the time has run out.

When the recognition engine 28 detects an utterance from the default grammar (i.e., the "yes" branch from step 210), the speech recognition user interface 30 provides three different affirmations that the system 20 heard and understood the command. The first affirmation is the emission of a distinctive "confirm" tone or sound from the audio output 110 or sound generator 158 (step 214). The speech recognition user interface 30 also briefly changes the color of the graphical display 300 from one color to another color to provide a visual affirmation that the command was correctly detected (step 216). The graphics are returned to their normal color after a preset duration (e.g., one-two seconds). The third affirmation is the resetting of the countdown graphic to its initial state, such as the full-length bar 310 in Fig. 7a (step 218).

At step 220 in Fig. 5, the speech recognition system determines whether the utterance represents a keyword in the application. If the user speaks the keyword such as "Place" (i.e., the "yes" branch from step 220), the speech recognition system temporarily expands the default grammar to include an additional set of utterances triggered by the keyword (step 222 in Fig. 5). In this case, the extra utterances are location-related terms, "City", "County", and "Street", as shown in grammar 254 of Fig. 6. Table 2 lists the ballooned grammar.

Table 2: Ballooned Grammar 254

<u>Word/Phrase</u>	<u>Command Function</u>
Place	Keyword to place menu
Name	Sets address book to specified name.
Directions	Keyword to Directions menu
Address	Keyword to address menu
City	Keyword to list of cities
County	Keyword to list of counties
Street	Keyword to list of streets

The process returns to the utterance detection step 210 in Fig. 5. Suppose that the user next speaks the word "City", which is also a keyword, within the prescribed response time. When the speech recognition engine 28 detects the keyword "City" (step 102 in Fig. 6), it emits a "confirm" tone, changes the graphics color, and restarts the countdown graphic (steps 214-218). Since "City" is also a keyword, the first ballooned grammar is further expanded to a second ballooned grammar 256 that adds the names of various U.S. cities, such as

1 “Albany”, “Anchorage”, “Atlanta”, “Baltimore”, “Boise”, “Boston”, “Buffalo”,
2 and so forth (step 222). The second ballooned grammar 256 thus contains the
3 original words in the default grammar 252, the added words in the first ballooned
4 grammar 254, and the new city names.

5 Suppose the user says the word “Atlanta” within the response time. The
6 speech recognition system recognizes the utterance “Atlanta” (i.e., the “yes”
7 branch from step 210), as represented by numeral 258 in Fig. 6. With this
8 recognition, the speech recognition UI 30 once again emits a “confirm” tone,
9 changes the graphics color, and restarts the countdown graphic. In this case,
10 however, the utterance is not a keyword (i.e., the “no” branch from step 220 in
11 Fig. 5). Accordingly, at step 224 in Fig. 5, the system performs the associated
12 function of filtering the address book to entries pertaining to the city Atlanta, as
13 represented by number 260 in Fig. 6.

14 After the associated operation is performed, the speech recognition system
15 returns to the small default grammar (step 226 in Fig. 5). This is represented
16 diagrammatically in Fig. 6 by the dashed return arrow to the default grammar 252.
17 The speech recognition user interface is advantageous because it provides both
18 auditory and visual feedback concerning the state of the speech recognition
19 system. The distinctive tones and countdown graphic informs the user as to
20 whether the system is awake or asleep, and if awake, how much time is left to
21 speak a command before the system goes dormant again. The visual portion is
22 designed to give quick information at a glance so that users, such as vehicle
23 drivers, need not concentrate for any significant time period to discern whether the
24 system is awaiting a response and there is still time to give one.
25

1 Although the invention has been described in language specific to structural
2 features and/or methodological steps, it is to be understood that the invention
3 defined in the appended claims is not necessarily limited to the specific features or
4 steps described. Rather, the specific features and steps are disclosed as preferred
5 forms of implementing the claimed invention.
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25

1 **CLAIMS**

2 1. A speech recognition system comprising:

3 a speech recognition engine to recognize an utterance, the speech
4 recognition engine being configured to actively listen for the utterance for a
5 predetermined response time; and

6 a user interface to provide visual and auditory feedback indicating whether
7 the speech recognition engine recognizes the utterance, the user interface being
8 configured to play an audible sound indicating recognition of the utterance and to
9 display a countdown graphic that changes with lapsing of the response time.

10
11 2. A speech recognition system as recited in claim 1, wherein the user
12 interface restarts the countdown graphic in the event the speech recognition engine
13 recognizes the utterance.

14
15 3. A speech recognition system as recited in claim 1, wherein the
16 response time is configurable.

17
18 4. A speech recognition system as recited in claim 1, wherein the user
19 interface displays visual elements in a first color and briefly changes to a second
20 color in the event the speech recognition engine recognizes the utterance.

21
22 5. A speech recognition system as recited in claim 1, wherein the
23 countdown graphic comprises a progress bar that shortens as the response time
24 diminishes.

1 6. A speech recognition system as recited in claim 1, wherein the user
2 interface plays another audible sound when the response time has elapsed.

3
4 7. A speech recognition system as recited in claim 1, wherein the speech
5 recognition engine is initially in a sleep mode and is awakened to an active mode
6 upon detection of a starter utterance, the user interface plays another audible sound
7 indicating that the speech recognition engine is in the active mode in the event the
8 speech recognition engine recognizes the starter utterance.

9
10 8. A speech recognition system as recited in claim 1, wherein the speech
11 recognition engine is initially in a sleep mode and is awakened to an active mode
12 upon depression of a button, the user interface plays another audible sound
13 indicating that the speech recognition engine is in the active mode in the event the
14 speech recognition engine recognizes the starter utterance.

15
16 9. A speech recognition system comprising:
17 an application;
18 a vocabulary accessible by the application, the vocabulary holding a set of
19 utterances applicable to the application;
20 a grammar that holds a subset of the utterances in the vocabulary;
21 a speech recognition engine to recognize the utterances in the grammar
22 within a predetermined response time; and
23 a user interface to display a countdown graphic that changes with lapsing of
24 the response time.
25

1 **10.** A speech recognition system as recited in claim 9, wherein the user
2 interface restarts the countdown graphic in the event the speech recognition engine
3 recognizes the one of the utterances.
4

5 **11.** A speech recognition system as recited in claim 9, wherein the user
6 interface displays visual elements in a first color and briefly changes to a second
7 color in the event the speech recognition engine recognizes one of the utterances.
8

9 **12.** A speech recognition system as recited in claim 9, wherein the
10 countdown graphic comprises a progress bar that shortens as the response time
11 diminishes.
12

13 **13.** A speech recognition system as recited in claim 9, wherein the user
14 interface plays an audible sound when the speech recognition engine recognizes
15 one of the utterances within the predetermine response time.
16

17 **14.** A speech recognition system as recited in claim 9, wherein the user
18 interface plays an audible sound when the response time has elapsed.
19

20 **15.** A speech recognition system as recited in claim 9, wherein the
21 speech recognition engine is initially in a sleep mode and is awakened to an active
22 mode upon detection of a starter utterance, the user interface plays another audible
23 sound indicating that the speech recognition engine is in the active mode in the
24 event the speech recognition engine recognizes the starter utterance.
25

1 **16.** An entertainment system incorporating the speech recognition
2 system as recited in claim 9.

3
4 **17.** A computing device incorporating the speech recognition system as
5 recited in claim 9.

6
7 **18.** A user interface for an speech recognition system, the user interface
8 comprising:

9 a display; and

10 a graphic progress bar shown on the display that indicates a response time
11 in which the speech recognition system is awaiting a user to speak, the progress
12 bar shortening with passage of the response time.

13
14 **19.** A user interface as recited in claim 18, wherein the graphic progress
15 bar is lengthened to its initial position after each recognized user input.

16
17 **20.** A user interface as recited in claim 18, wherein the graphic progress
18 bar briefly changes color when a user input is recognized.

19
20 **21.** A speech recognition system incorporating the user interface as
21 recited in claim 18.

22
23 **22.** A computing device incorporating the user interface as recited in
24 claim 18.

1 **23.** A user interface for an speech recognition system, the user interface
2 comprising:

3 a display;

4 an audio input to receive audible utterances;

5 a graphic shown on the display that indicates a response time in which the
6 speech recognition system is awaiting receipt of an utterance via the audio input;
7 and

8 an audio generator to emit an audible sound when the speech recognition
9 system recognizes the utterance.

10
11 **24.** A user interface as recited in claim 23, wherein the audio generator
12 emits a second audible sound when the speech recognition system fails to
13 recognize the utterance within the response time.

14
15 **25.** A speech recognition system incorporating the user interface as
16 recited in claim 23.

17
18 **26.** A computing device incorporating the user interface as recited in
19 claim 23.

20
21 **27.** A vehicle computer system comprising:
22 a computer;
23 an open platform operating system executing on the computer, the
24 operating system being configured to support multiple applications; and
25

1 a speech recognition system to detect utterances used to control at least one
2 of the applications running on the computer, the speech recognition system having
3 a user interface to provide visual and auditory feedback indicating whether an
4 utterance is recognized, the user interface being configured to play an audible
5 sound indicating recognition of the utterance and to display a graphic that changes
6 with passage of the response time.

7
8 **28.** A vehicle computer system as recited in claim 27, wherein the user
9 interface restarts the graphic in the event the utterance is recognized.

10
11 **29.** A vehicle computer system as recited in claim 27, wherein the user
12 interface displays visual elements in a first color and briefly changes to a second
13 color in the event the utterance is recognized.

14
15 **30.** A vehicle computer system as recited in claim 27, wherein the
16 graphic comprises a progress bar that shortens as the response time passes.

17
18 **31.** A vehicle computer system as recited in claim 27, wherein the user
19 interface plays another audible sound when the response time has elapsed.

1 **32.** A vehicle computer system as recited in claim 27, wherein the
2 speech recognition system is initially in a sleep mode and is awakened to an active
3 mode upon detection of a starter utterance, the user interface plays another audible
4 sound indicating that the speech recognition system is in the active mode in the
5 event the starter utterance is recognized.

6
7 **33.** A collaboration system involving multiple interconnected devices
8 comprising:

9 a voice input mechanism resident at each of the devices;
10 an audio output system resident at each of the devices; and
11 a user interface to provide visual and auditory feedback indicating when a
12 party located at one of the devices can speak, the user interface being configured
13 to play an audible sound when the party can begin speaking and to display a
14 graphic that changes with lapsing of time to indicate a duration that the party can
15 speak.

16
17 **34.** A method for operating a speech recognition system, comprising the
18 following steps:

19 initiating a response time in which to receive an audible utterance;
20 displaying a graphic representing the response time; and
21 changing the graphic to indicate passage of the response time.

1 **35.** A method as recited in claim 34, wherein the displaying and
2 changing steps comprise the steps of depicting a progress bar and shortening the
3 progress bar as the response time passes.

4
5 **36.** A method as recited in claim 34, further comprising the step of
6 resetting the graphic when an audible utterance is recognized.

7
8 **37.** A method as recited in claim 34, further comprising the step of
9 changing a color of the graphic when an audible utterance is recognized.

10
11 **38.** A method as recited in claim 34, further comprising the step of
12 playing a sound when an audible utterance is recognized.

13
14 **39.** A method as recited in claim 34, further comprising the step of
15 playing a sound when no audible utterance is recognized within the response time.
16
17
18
19
20
21
22
23
24
25

1 **ABSTRACT**

2 A speech recognition system having a user interface that provides both
3 visual and auditory feedback to a user. The user interface includes an audio sound
4 or speech generator that produces three distinct sounds: an "on" sound signifying
5 that the speech recognition system is on and actively awaiting vocal input; an
6 "off" sound indicating that the speech recognition system is off and in a sleep
7 mode; and a "confirm" sound noting that an utterance has been recognized. The
8 "on" sound is triggered by a key "wake up" command or by depression of button.
9 Once awake, the speech recognition engine expects to receive an utterance within
10 a predetermined response time. The "confirm" sound signals the start of the
11 response time. If the response time lapses before a recognizable utterance is
12 entered, the "off" sound is played. The user interface further includes a visual
13 component in the form of a graphic that changes with the tolling of the response
14 period. In one implementation, the count graphic is a progress bar that counts
15 down or shortens in proportion to the passage of the response period. When the
16 response time runs out, the progress bar disappears entirely. On the other hand, if
17 the speech engine recognizes an utterance within the response period, the user
18 interface plays the "confirm" sound and restarts the countdown graphic. The user
19 interface may also change the color of the graphic elements briefly to reflect a
20 correct voice entry.
21
22
23
24
25

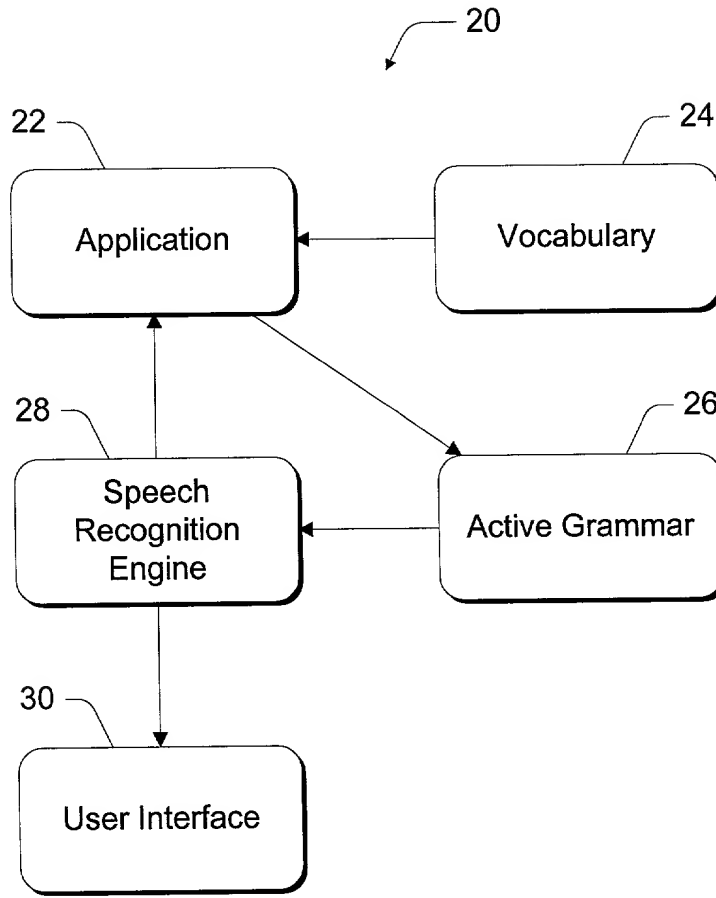
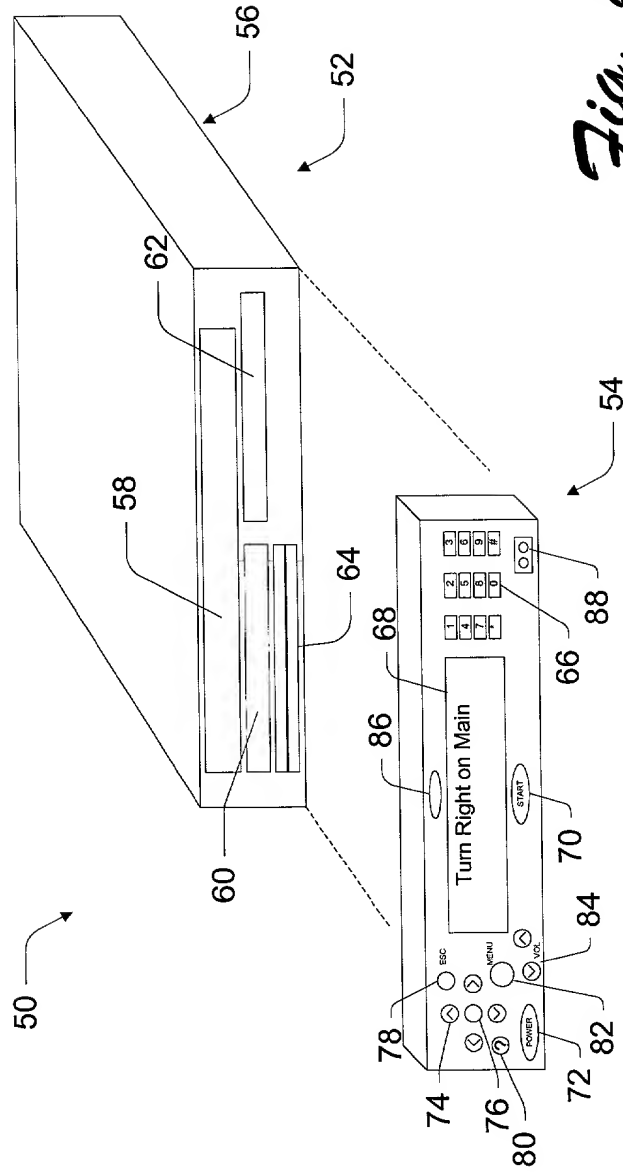


Fig. 1



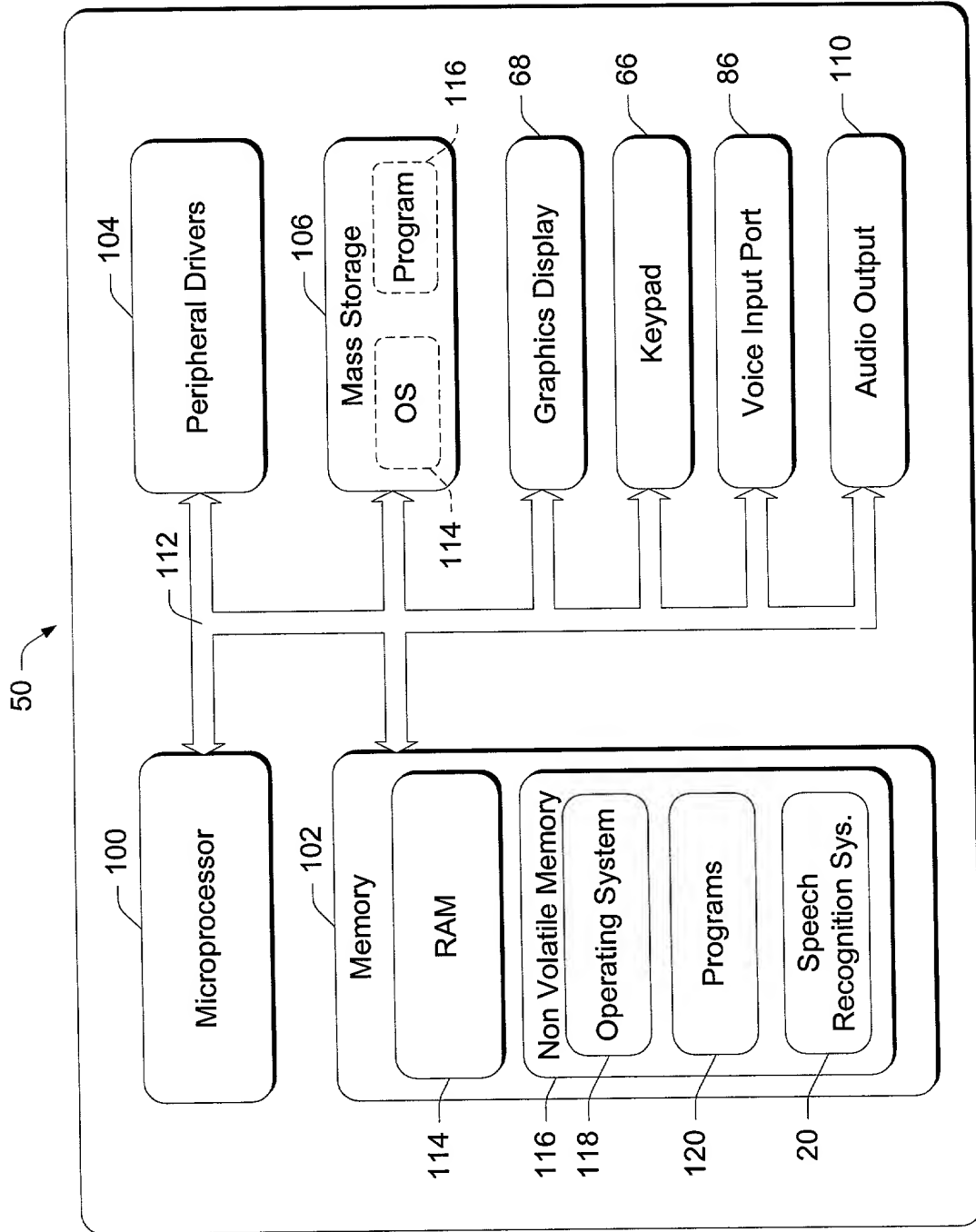


Fig. 3

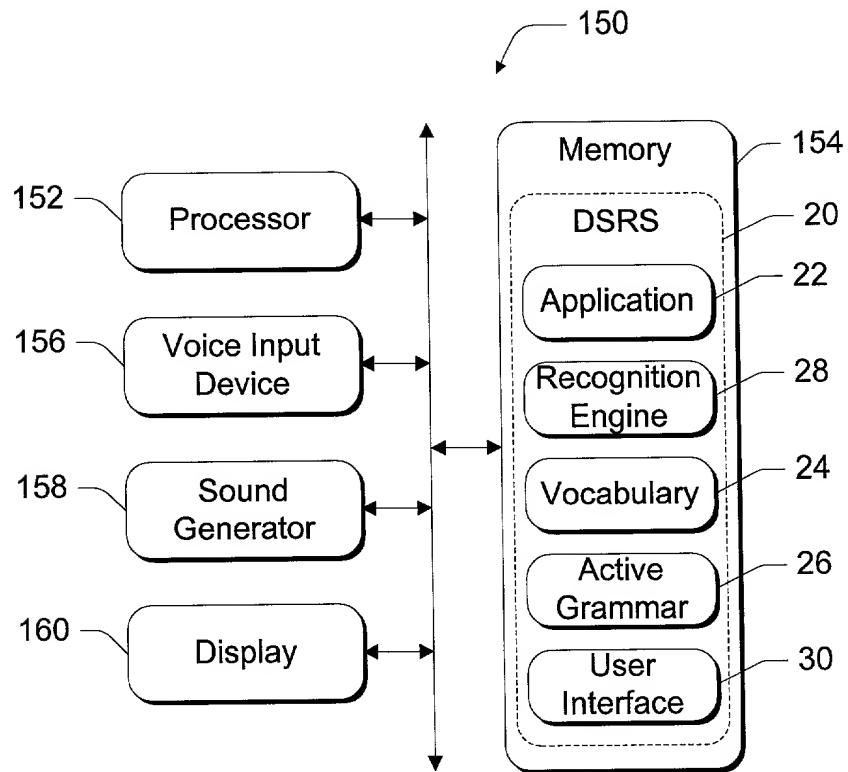


Fig. 4

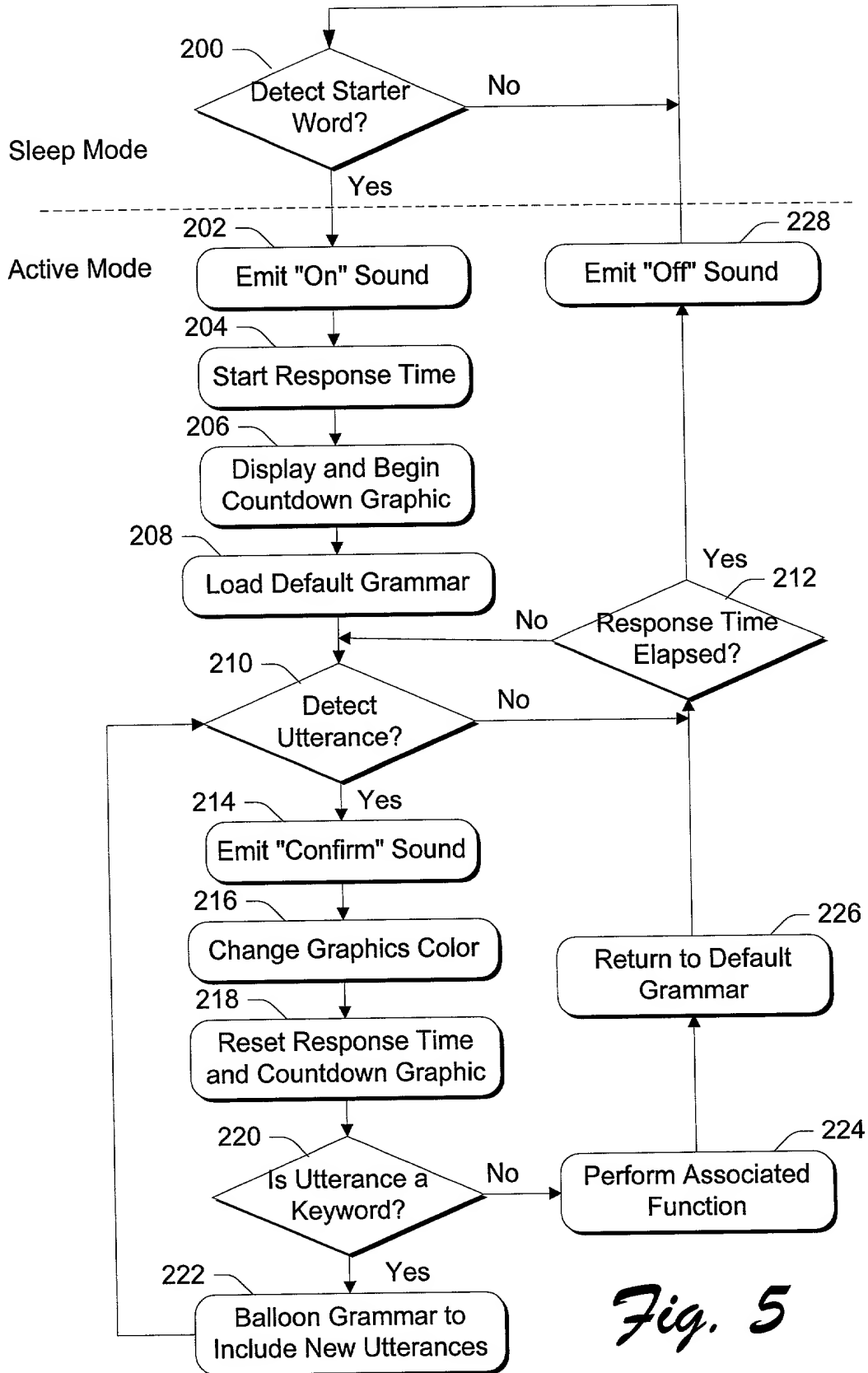


Fig. 5

Patent 2016060

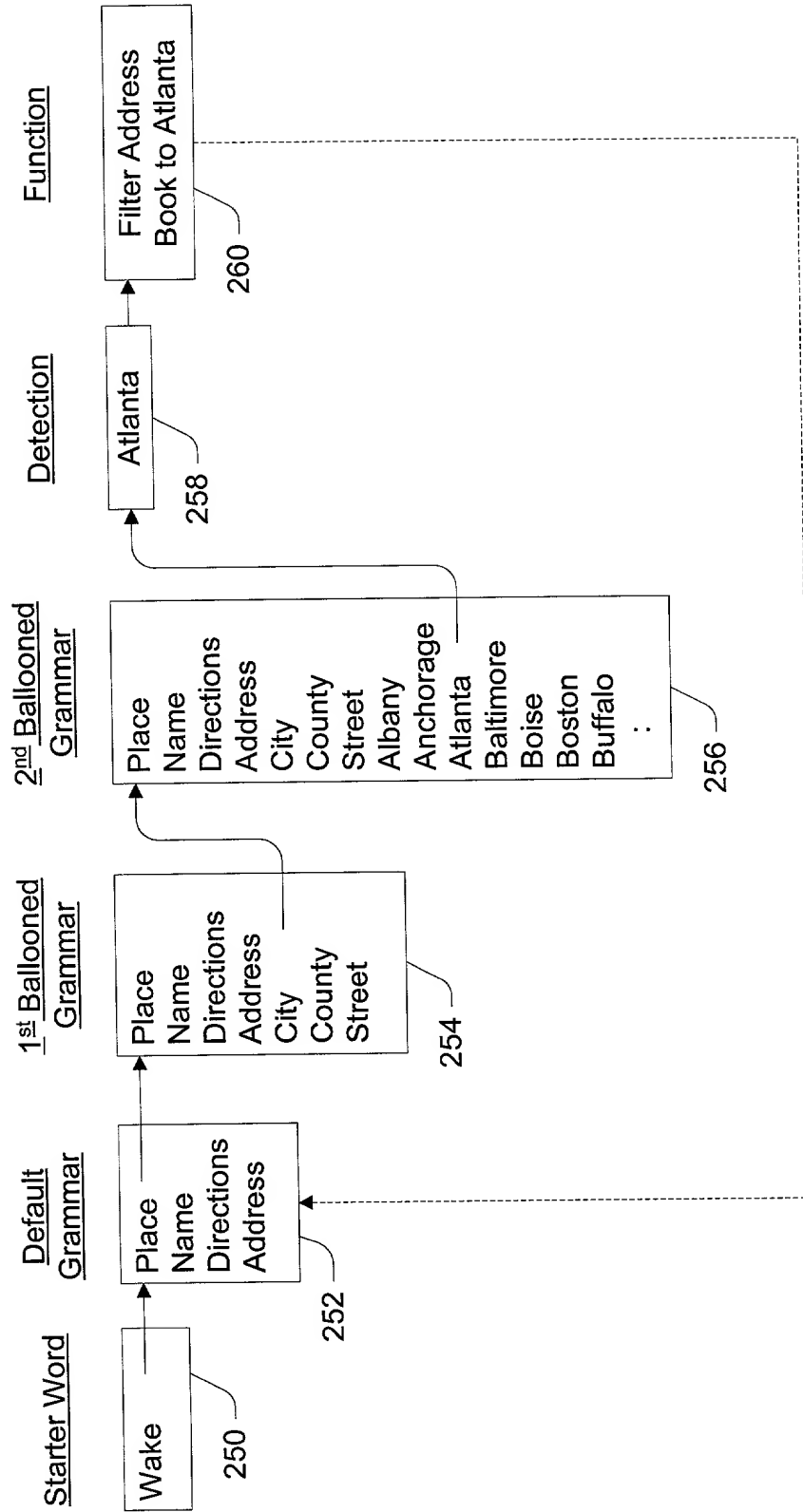


Fig. 6

Fig. 7a

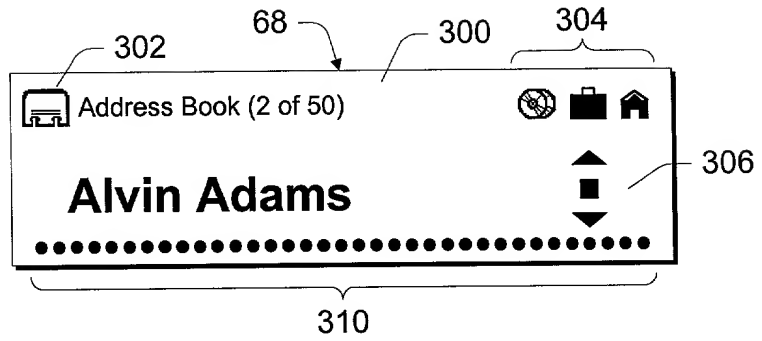


Fig. 7b

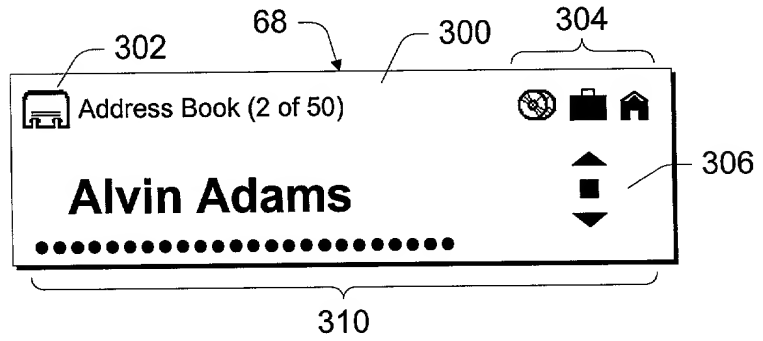
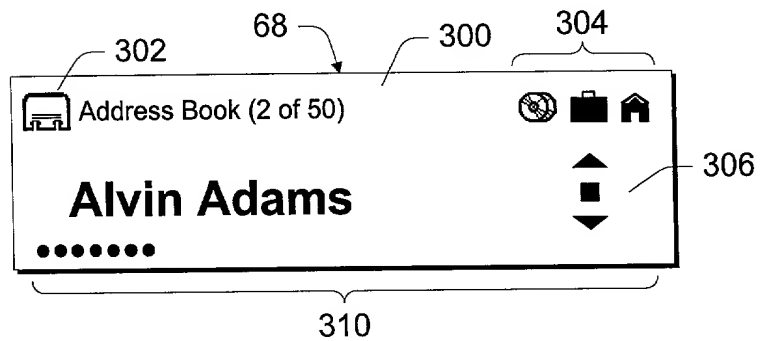


Fig. 7c



1 **IN THE UNITED STATES PATENT AND TRADEMARK OFFICE**

2 Inventorship..... Sarah E. Zuberec et al.
 3 Applicant Microsoft Corporation
 4 Attorney's Docket No. MS1-286US
 5 Title:..... Speech Recognition User Interface

6 **DECLARATION FOR PATENT APPLICATION**

7 As a below named inventor, I hereby declare that:

8 My residence, post office address and citizenship are as stated below next to
 9 my name.

10 I believe I am the original, first and sole inventor (if only one name is listed
 11 below) or an original, first and joint inventor (if plural names are listed below) of the
 12 subject matter which is claimed and for which a patent is sought on the invention
 13 entitled "Speech Recognition User Interface" the specification of which is attached
 14 hereto.

15 I have reviewed and understand the content of the above-identified
 16 specification, including the claims.

17 I acknowledge the duty to disclose information which is material to the
 18 examination of this application in accordance with Title 37, Code of Federal
 19 Regulations, § 1.56(a).

20 PRIOR FOREIGN APPLICATIONS: no applications for foreign patents or
 21 inventor's certificates have been filed prior to the date of execution of this
 22 declaration.

Power of Attorney

I appoint the following attorneys to prosecute this application and transact all future business in the Patent and Trademark Office connected with this application: Lewis C. Lee, Reg. No. 34,656; Daniel L. Hayes, Reg. No. 34,618; David A. Morach, Reg. No. 42,905 Katie E. Sako, Reg. No. 32,628 and Daniel D. Crouse, Reg. No. 32,022.

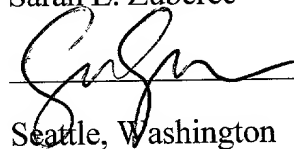
Send correspondence to: LEE & HAYES, PLLC, W. 201 North River Drive, Suite 430, Spokane, Washington, 99201. Direct telephone calls to: Daniel L. Hayes (509) 324-9256.

All statements made herein of my own knowledge are true and that all statements made on information and belief are believed to be true; and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code and that such willful false statement may jeopardize the validity of the application or any patent issued therefrom.

Full name of inventor:

Sarah E. Zuberec

Inventor's Signature



Date: Nov 6/98-

Residence:

Seattle, Washington

Citizenship:

Canada

Post Office Address:

1631 16th Ave. #307
Seattle, WA 98122

Full name of inventor:

Cynthia DuVal

Inventor's Signature

Cynthia DuVal Date: 11/9/98

Residence:

Shoreline, Washington

Citizenship:

USA

Post Office Address:

1518 N. 146th Street
Shoreline, WA 98133

Full name of inventor:

Benjamin N. Rabelos

Inventor's Signature

B. N. Rabelos Date: 11/6/98

Residence:

Seattle, Washington

Citizenship:

USA

Post Office Address:

3854 30th Avenue W.
Seattle, WA 98199